

AutoBayes—Automatic Synthesis of Statistical Data Analysis Programs from Bayesian Networks

Bernd Fischer, Tom Pressburger, Johann Schumann

The goal of the AutoBayes project is to make statistical data analysis easier and more accessible to scientists by automatically synthesizing efficient data-analysis programs from statistical models that are used for the definition of valid information. Data analysis can be defined as any process that extracts more abstract information from mere data. It includes such diverse tasks as general parameter estimation and curve fitting, clustering and classification, data compression, fusion of heterogeneous data sources, change and anomaly detection in time-series or image data, or image segmentation. Although there are many approaches to data analysis, statistical data analysis is the only mathematically rigorous approach. In statistical data analysis, a statistical model is used to define how much information the data originally contains, and thus, how much statistically valid information can ultimately be extracted from the data. This approach is standard in medical sciences such as epidemiology, where the cost of wrong conclusions can be high, and statistical data analysis is now becoming more widespread within the fields relevant to NASA. Unfortunately, the development of statistical data-analysis programs is expensive and time-consuming, and it requires expertise at the intersection of computer science, statistics, and the application.

AutoBayes uses a notation based on Bayesian networks to specify the statistical model. A Bayesian network is essentially a directed, acyclic graph where the nodes represent random variables, that is, the known (that is, data) or unknown (that is, model parameters) objects of interest, and the edges represent conditional dependency or influence. The network thus encodes the full joint probability distribution over the model. This study has extended the standard notation to handle identically distributed random variables (for example, sensor arrays or image pixels) more efficiently.

Figure 1 illustrates a simple but typical data-analysis problem, a so-called "mixture problem": given the data points, which are specified to have been generated from several Gaussian distributions

with unknown mean and variance, summarize each cluster by its most likely mean and variance.

AutoBayes generates programs that apply known statistical techniques, for example, "Expectation Maximization," to solve the given statistical problems. For each technique, the AutoBayes program knows the conditions under which the technique is applicable, the class of data-analysis problems the technique solves, and how the technique is applied to a specific problem. The generation process takes place in the framework of automated theorem proving, guided by the applicability conditions of the statistical techniques. The theorem-proving framework lends some assurance that the generated program is correct.

FY99 saw the completion of the first prototype of the AutoBayes synthesis system. AutoBayes extracts the network from a textual specification language (see figure 2 for a possible model of the mixture problem) and currently generates code in Matlab or C++; other target languages can be added easily. It has been tested on several textbook examples. Researchers expect to apply it to an image-processing problem arising from a project searching for planets around other stars by looking for the dip in brightness as the planet transits in front of the star.

Point of Contact: B. Fischer
(650) 604-2977
fisch@ptolemy.arc.nasa.gov

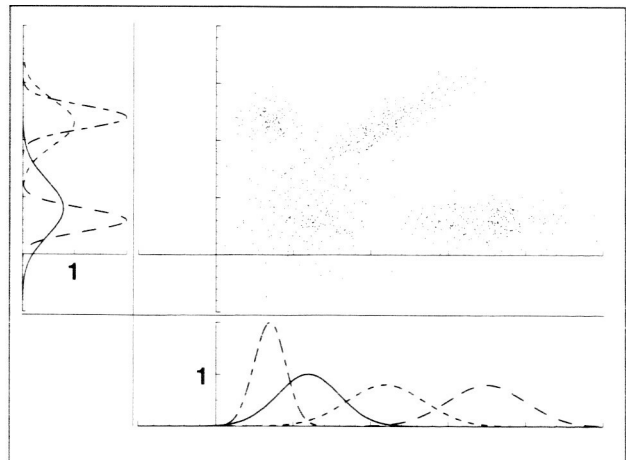


Fig. 1. Example problem—mixture of Gaussians.

```

constant n_points;
constant n_classes;

probability rho[n_classes];

random mu_x[n_classes];
random mu_y[n_classes];
random sigma_x[n_classes];
random sigma_y[n_classes];

random c[n_points];

data x[n_points]
data y[n_points]

c ~ discrete(rho)

x[i] ~ gaussian(mu_x[c[i]], sigma_x[c[i]]);
y[i] ~ gaussian(mu_y[c[i]], sigma_y[c[i]]);

optimize P(x, y | rho, mu_x, mu_y, sigma_x, sigma_y)
for rho, mu_x, mu_y, sigma_x, sigma_y;

```

Fig. 2. Example specifications.

Propellant Preservation for Mars Missions

Peter Kittel

The last few years have seen extensive technology planning for human missions to Mars. These missions will make extensive use of cryogenic propellants, some of which will be transported to Mars. Additional propellants will be manufactured, liquefied, and stored on Mars. The missions that use these propellants could start early this century. Although many of the plans are still evolving, it has been possible to derive a set of cooler requirements. Recent estimates of these requirements are given here along with a discussion of whether the requirements can be met with existing coolers and coolers currently being developed.

In recent years, a variety of transportation scenarios have been considered. The analysis reported here relates to one of the more promising nonnuclear options. It makes use of solar electric propulsion (SEP). The SEP baseline concept follows:

1. A SEP tug boosts the TMI (trans-Mars injection) stage to a highly elliptical orbit. This phase requires

400 days of propellant storage for the TMI piloted stage and 250 days for the TMI cargo stages.

2. The ascent and descent stages require about 580 days of propellant storage for the piloted mission and 550 days for the cargo case. These stages use oxygen (O₂) and methane (CH₄) for propellants.

3. The TEI (trans-Earth injection) stage requires a storage duration of 1200 days. This stage uses hydrogen (H₂) as the propellant.

4. All tanks are cooled by cryocoolers to eliminate boil off.

5. The tank design is standardized to 3.29-meter-diameter spheres.

Fixing the volume results in the stages having multiple tanks, with many of the tanks full at launch.

A thermal model was used to estimate the cooler requirements. This model could estimate the tank size and mass for MLI (multilayer insulation) insulated tanks with and without coolers. The cooling power and the mass of the power source and radiators were included for ZBO (zero boil-off) storage. In the first case, the volume of the tank was variable. The volume was adjusted until it was large enough to accommodate the boil-off during the mission and still preserve the required propellant until it was needed. In the latter case, the volume was fixed. An optional cooled shield has been included in the model for hydrogen tanks.

The results of the model are presented in Table 1. Pulse tube coolers that can meet these requirements are being developed.

Point of Contact: P. Kittel
(650) 604-4297
pkittel@mail.arc.nasa.gov

Table 1. Minimum set of coolers assuming 250 kelvin heat rejection temperature

Cooler	Cooler Power (Watts)	Temperature (kelvin)
Single stage	11.8	97.2
Two-stage		
First stage	18.6	85
Second-stage	1.2	22.8